



PCT

## RAW SEQUENCE LISTING

DATE: 12/11/2002

PATENT APPLICATION: US/09/913,494A

TIME: 09:37:39

Input Set : N:\Crf4\12022002\I913494.raw  
 Output Set: N:\CRF4\12112002\I913494A.raw

1 <110> APPLICANT: Merck Patent GmbH  
 2 <120> TITLE OF INVENTION: Glucose Dehydrogenase Fusion Proteins and their Utilization  
 in Expression

3 Systems  
 4 <130> FILE REFERENCE: Merck 2289 / P9906920  
 5 <140> CURRENT APPLICATION NUMBER: US/09/913,494A  
 C--> 6 <141> CURRENT FILING DATE: 1999-02-19  
 7 <150> PRIOR APPLICATION NUMBER: DE 19906920  
 8 <151> PRIOR FILING DATE: 1999-02-19  
 9 <160> NUMBER OF SEQ ID NOS: 17  
 10 <170> SOFTWARE: PatentIn version 3.1  
 12 <210> SEQ ID NO: 1  
 13 <211> LENGTH: 3992  
 14 <212> TYPE: DNA  
 15 <213> ORGANISM: Bacillus megaterium  
 16 <220> FEATURE:  
 17 <221> NAME/KEY: CDS  
 18 <222> LOCATION: (186)..(968)  
 19 <223> OTHER INFORMATION: Glucose Dehydrogenase from Bacillus megaterium  
 20 <220> FEATURE:  
 21 <221> NAME/KEY: gene  
 22 <222> LOCATION: (1)..(3992)  
 23 <223> OTHER INFORMATION: plasmid PAW2  
 24 <220> FEATURE:  
 25 <221> NAME/KEY: CDS  
 26 <222> LOCATION: (978)..(1010)  
 27 <223> OTHER INFORMATION: poly-histidin tag  
 28 <400> SEQUENCE: 1

ENTERED

```

29 ccacgcaatg gccagatgat taattcctaa tttttgttga cactctatca ttgatagagt      60
30 tattttacca ctccctatca gtgatagaga aaagtgaat gaatagttcg acaaaaatct      120
31 agataacgag ggcaatcgat gaattcgagc tcggtaccgc gggatccctc gaggtcgacc      180
32 tgcag atg tat aca gat tta aaa gat aaa gta gtt gta att aca ggt gga      230
33   Met Tyr Thr Asp Leu Lys Asp Lys Val Val Val Ile Thr Gly Gly
34     1             5             10             15
35 tca aca ggt tta gga cgc gca atg gct gtt cgt ttc ggt caa gaa gaa      278
36 Ser Thr Gly Leu Gly Arg Ala Met Ala Val Arg Phe Gly Gln Glu Glu
37           20             25             30
38 gca aaa gtt gtt att aac tat tac aac aat gaa gaa gaa gct cta gat      326
39 Ala Lys Val Val Ile Asn Tyr Tyr Asn Asn Glu Glu Glu Ala Leu Asp
40           35             40             45
41 gcg aaa aaa gaa gta gaa gaa gca ggc gga caa gca atc atc gtt caa      374
42 Ala Lys Lys Glu Val Glu Glu Ala Gly Gly Gln Ala Ile Ile Val Gln
43           50             55             60
44 ggc gat gta aca aaa gaa gaa gac gtt gta aat ctt gtt caa aca gct      422

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/913,494A

DATE: 12/11/2002

TIME: 09:37:39

Input Set : N:\Cr4\12022002\I913494.raw

Output Set: N:\CRF4\12112002\I913494A.raw

45	Gly Asp Val Thr Lys Glu Glu Asp Val Val Asn Leu Val Gln Thr Ala	
46	65 70 75	
47	att aaa gaa ttt ggt aca tta gac gta atg att aac aac gct ggt gtt	470
48	Ile Lys Glu Phe Gly Thr Leu Asp Val Met Ile Asn Asn Ala Gly Val	
49	80 85 90 95	
50	gaa aac cca gtt cct tct cat gag cta tct cta gat aac tgg aac aaa	518
51	Glu Asn Pro Val Pro Ser His Glu Leu Ser Leu Asp Asn Trp Asn Lys	
52	100 105 110	
53	gtt att gat aca aac tta aca ggt gca ttc tta gga agc cgt gaa gca	566
54	Val Ile Asp Thr Asn Leu Thr Gly Ala Phe Leu Gly Ser Arg Glu Ala	
55	115 120 125	
56	att aaa tac ttc gtt gaa aac gac att aaa gga aat gtt atc aac atg	614
57	Ile Lys Tyr Phe Val Glu Asn Asp Ile Lys Gly Asn Val Ile Asn Met	
58	130 135 140	
59	tct agc gtt cac gaa atg att cct tgg cca tta ttt gtt cac tac gca	662
60	Ser Ser Val His Glu Met Ile Pro Trp Pro Leu Phe Val His Tyr Ala	
61	145 150 155	
62	gca agt aaa ggc ggt atg aaa cta atg acg gaa aca ttg gct ctt gaa	710
63	Ala Ser Lys Gly Gly Met Lys Leu Met Thr Glu Thr Leu Ala Leu Glu	
64	160 165 170 175	
65	tat gcg cca aaa ggt atc cgc gta aat aat att gga cca ggt gcg atg	758
66	Tyr Ala Pro Lys Gly Ile Arg Val Asn Asn Ile Gly Pro Gly Ala Met	
67	180 185 190	
68	aac aca cca att aac gca gag aaa ttt gca gat cca gaa caa cgt gca	806
69	Asn Thr Pro Ile Asn Ala Glu Lys Phe Ala Asp Pro Glu Gln Arg Ala	
70	195 200 205	
71	gac gta gaa agc atg att cca atg ggt tac atc ggt aaa cca gaa gaa	854
72	Asp Val Glu Ser Met Ile Pro Met Gly Tyr Ile Gly Lys Pro Glu Glu	
73	210 215 220	
74	gta gca gca gtt gca gca ttc tta gct tca tca caa gca agc tat gta	902
75	Val Ala Ala Val Ala Ala Phe Leu Ala Ser Ser Gln Ala Ser Tyr Val	
76	225 230 235	
77	aca ggt att aca tta ttt gca gat ggc ggt atg acg aaa tac cct tct	950
78	Thr Gly Ile Thr Leu Phe Ala Asp Gly Gly Met Thr Lys Tyr Pro Ser	
79	240 245 250 255	
80	ttc caa gca gga aga ggc taatagagc gct atg aga gga tcg cat cac cat	1001
81	Phe Gln Ala Gly Arg Gly Ala Met Arg Gly Ser His His His	
82	260 265	
83	cac cat cac taatagaagc ttgacctgtg aagtgaaaaa tggcgccat	1050
84	His His His	
85	270	
86	tgtgcgacat tttttttgtc tgccggtttac cgctactgcg tcacggatct ccacgcgccc	1110
87	tgtagcggcg cattaagcgc ggcggtgtg gtggttacgc gcagcgtgac cgctacactt	1170
88	gccagcgcgc tagcgcgcgc tcctttcgct ttcttcctt cctttctcgc cacgttcgcc	1230
89	ggctttcccc gtcaagctct aaatcgggg ctcccttttag ggttcggatt tagtgcttta	1290
90	cggcacctcg acccaaaaa acttgattag ggtgatggtt cacgtagtg gcatcgccc	1350
91	tgatagacgg tttttcgccc ttgacgttg gagtccacgt tctttaatag tggactcttg	1410
92	ttccaaactg gaacaacact caacctatc tcggtctatt cttttgattt ataagggtt	1470
93	ttgccgattt cggcctattg gttaaaaaat gagctgattt aacaaaaatt taacgcgaat	1530

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/913,494A

DATE: 12/11/2002

TIME: 09:37:39

Input Set : N:\Crf4\12022002\I913494.raw

Output Set: N:\CRF4\12112002\I913494A.raw

94	tttaacaaaa	tattaacgct	tacaatttca	ggtggcactt	ttcgggggaaa	tgtgcgcgga	1590
95	accctatatt	gtttattttt	ctaaatacat	tcaaataatgt	atccgctcat	gagacaataa	1650
96	ccctgataaa	tgcttcaata	atattgaaaa	aggaagagta	tgagtattca	acatttccgt	1710
97	gtcgccctta	ttcccttttt	tgcggcattt	tgcttctctg	tttttgctca	cccagaaaacg	1770
98	ctgggtgaaag	taaaagatgc	tgaagatcag	ttgggtgcac	gagtgggtta	catcgaactg	1830
99	gatctcaaca	gcggtgaagat	ccttgagagt	tttcgccccg	aagaacgttt	tccaatgatg	1890
100	agcactttta	aagttctgct	atgtggcgcg	gtattatccc	gtattgacgc	cgggcaagag	1950
101	caactcggtc	gccgcataca	ctattctcag	aatgacttgg	ttgagtactc	accagtcaca	2010
102	gaaaagcatc	ttacggatgg	catgacagta	agagaattat	gcagtgtctc	cataaccatg	2070
103	agtataaca	ctgcggccaa	cttacttctg	acaacgatcg	gaggaccgaa	ggagctaacc	2130
104	gcttttttgc	acaacatggg	ggatcatgta	actcgccttg	atcggtggga	accggagctg	2190
105	aatgaagcca	taccaaacga	cgagcgtgac	accacgatgc	ctgtagcaat	ggcaacaacg	2250
106	ttgcgcaaac	tattaactgg	cgaactactt	actctagctt	cccggcaaca	attgatagac	2310
107	tggatggagg	cggataaagt	tgcaggacca	cttctgcgct	cggcccttcc	ggctggctgg	2370
108	tttattgctg	ataaatctgg	agccgggtgag	cgtggctctc	gcggtatcat	tgcagcactg	2430
109	gggccaagat	gtaagccctc	ccgtatcgta	gttatctaca	cgacggggag	tcaggcaact	2490
110	atggatgaac	gaaatagaca	gatcgctgag	ataggtgcct	cactgattaa	gcattggtag	2550
111	gaattaatga	tgtctcgttt	agataaaaagt	aaagtgatta	acagcgcat	agagctgctt	2610
112	aatgaggtcg	gaatcgaagg	tttaacaacc	cgtaaactcg	cccagaagct	aggtgtagag	2670
113	cagcctacat	tgtattggca	tgtaaaaaat	aagcgggctt	tgctcgacgc	cttagccatt	2730
114	gagatgttag	ataggcacca	tactcacttt	tgcccttttag	aaggggaaaag	ctggcaagat	2790
115	tttttacgta	ataacgctaa	aagttttaga	tgtgctttac	taagtcatcg	cgatggagca	2850
116	aaagtacatt	taggtacacg	gcctacagaa	aaacagtatg	aaactctcga	aaatcaatta	2910
117	gcctttttat	gccaacaagg	tttttacta	gagaatgcat	tatatgcact	cagcgcagtg	2970
118	gggcatttta	ctttaggttg	cgtattggaa	gatcaagagc	atcaagtgcg	taaaagaagaa	3030
119	agggaaacac	ctactactga	tagtatgccg	ccattattac	gacaagctat	cgaattattt	3090
120	gatcaccaag	gtgcagagcc	agccttctta	ttcggccttg	aattgatcat	atgcggatta	3150
121	gaaaaacaac	ttaaagtgtg	aagtgggtct	taaaagcagc	ataacctttt	tccgtgatgg	3210
122	taacttcact	agttttaaag	gatctagggt	aagatccctt	ttgataatct	catgaccaaa	3270
123	atcccttaac	gtgagttttc	gttccactga	gcgtcagacc	ccgtagaaaa	gatcaaagga	3330
124	tcttcttgag	atcctttttt	tctgcgcgta	atctgctgct	tgcaaacaaa	aaaaccaccg	3390
125	ctaccagcgg	tggtttgttt	gccggatcaa	gagctaccaa	ctctttttcc	gaaggtaact	3450
126	ggcttcagca	gagcgcagat	accaaatact	gtccttctag	tgtagccgta	gttaggccac	3510
127	cacttcaaga	actctgtagc	accgcctaca	tacctcgctc	tgctaatact	gttaccagtg	3570
128	gctgctgcca	gtggcgataa	gtcgtgtctt	accgggttgg	actcaagacg	atagttaccg	3630
129	gataaggcgc	agcggtcggg	ctgaacgggg	ggttcgtgca	cacagcccag	cttgagcgca	3690
130	acgacctaca	ccgaactgag	atacctacag	cgtgagctat	gagaaaagcg	cacgcttccc	3750
131	gaaggagagaa	aggcggacag	gtatccggta	agcggcaggg	tcggaacagg	agagcgcacg	3810
132	agggagcttc	cagggggaaa	cgcttggtat	ctttatagtc	ctgtcgggtt	tcgccacctc	3870
133	tgacttgagc	gtcgattttt	gtgatgctcg	tcaggggggc	ggagcctatg	gaaaaacgcc	3930
134	agcaacgcgg	ccttttttacg	gttcctggcc	ttttgctggc	cttttgctca	catgacccca	3990
135	ca						3992
137	<210>	SEQ ID NO: 2					
138	<211>	LENGTH: 261					
139	<212>	TYPE: PRT					
140	<213>	ORGANISM: Bacillus megaterium					
141	<400>	SEQUENCE: 2					
142		Met Tyr Thr Asp Leu Lys Asp Lys Val Val Val Ile Thr Gly Gly Ser					
143		1 5 10 15					

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/913,494A

DATE: 12/11/2002

TIME: 09:37:39

Input Set : N:\Crf4\12022002\I913494.raw

Output Set: N:\CRF4\12112002\I913494A.raw

```

144 Thr Gly Leu Gly Arg Ala Met Ala Val Arg Phe Gly Gln Glu Glu Ala
145          20          25          30
146 Lys Val Val Ile Asn Tyr Tyr Asn Asn Glu Glu Glu Ala Leu Asp Ala
147          35          40          45
148 Lys Lys Glu Val Glu Glu Ala Gly Gly Gln Ala Ile Ile Val Gln Gly
149          50          55          60
150 Asp Val Thr Lys Glu Glu Asp Val Val Asn Leu Val Gln Thr Ala Ile
151          65          70          75          80
152 Lys Glu Phe Gly Thr Leu Asp Val Met Ile Asn Asn Ala Gly Val Glu
153          85          90          95
154 Asn Pro Val Pro Ser His Glu Leu Ser Leu Asp Asn Trp Asn Lys Val
155          100          105          110
156 Ile Asp Thr Asn Leu Thr Gly Ala Phe Leu Gly Ser Arg Glu Ala Ile
157          115          120          125
158 Lys Tyr Phe Val Glu Asn Asp Ile Lys Gly Asn Val Ile Asn Met Ser
159          130          135          140
160 Ser Val His Glu Met Ile Pro Trp Pro Leu Phe Val His Tyr Ala Ala
161          145          150          155          160
162 Ser Lys Gly Gly Met Lys Leu Met Thr Glu Thr Leu Ala Leu Glu Tyr
163          165          170          175
164 Ala Pro Lys Gly Ile Arg Val Asn Asn Ile Gly Pro Gly Ala Met Asn
165          180          185          190
166 Thr Pro Ile Asn Ala Glu Lys Phe Ala Asp Pro Glu Gln Arg Ala Asp
167          195          200          205
168 Val Glu Ser Met Ile Pro Met Gly Tyr Ile Gly Lys Pro Glu Glu Val
169          210          215          220
170 Ala Ala Val Ala Ala Phe Leu Ala Ser Ser Gln Ala Ser Tyr Val Thr
171          225          230          235          240
172 Gly Ile Thr Leu Phe Ala Asp Gly Gly Met Thr Lys Tyr Pro Ser Phe
173          245          250          255
174 Gln Ala Gly Arg Gly
175          260
177 <210> SEQ ID NO: 3
178 <211> LENGTH: 11
179 <212> TYPE: PRT
180 <213> ORGANISM: Bacillus megaterium
181 <400> SEQUENCE: 3
182 Ala Met Arg Gly Ser His His His His His His
183 1 5 10
185 <210> SEQ ID NO: 4
186 <211> LENGTH: 4193
187 <212> TYPE: DNA
188 <213> ORGANISM: Bacillus megaterium / Heamenteria ghiliani fusion
189 <220> FEATURE:
190 <221> NAME/KEY: CDS
191 <222> LOCATION: (141)..(344)
192 <223> OTHER INFORMATION: Tridegin
193 <220> FEATURE:
194 <221> NAME/KEY: gene

```

## RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/913,494A

DATE: 12/11/2002

TIME: 09:37:39

Input Set : N:\CrF4\12022002\I913494.raw

Output Set: N:\CRF4\12112002\I913494A.raw

```

195 <222> LOCATION: (1)..(4193)
196 <223> OTHER INFORMATION: plasmid pAW4
197 <220> FEATURE:
198 <221> NAME/KEY: CDS
199 <222> LOCATION: (387)..(1169)
200 <223> OTHER INFORMATION: Glucose Dehydrogenase
201 <220> FEATURE:
202 <221> NAME/KEY: CDS
203 <222> LOCATION: (1179)..(1211)
204 <223> OTHER INFORMATION: poly histidine tag
205 <400> SEQUENCE: 4
206 ccacgcaatg gccagatgat taattcctaa tttttgttga cactctatca ttgatagagt 60
207 tattttacca ctccctatca gtgatagaga aaagtgaat gaatagttcg acaaaaatct 120
208 agataacgag ggcaatcgat atg aaa cta ttg cct tgc aaa gaa tgg cat caa 173
209 Met Lys Leu Leu Pro Cys Lys Glu Trp His Gln
210 1 5 10
211 ggt att cct aac cct agg tgc tgg tgt ggg gct gat cta gaa tgc gca 221
212 Gly Ile Pro Asn Pro Arg Cys Trp Cys Gly Ala Asp Leu Glu Cys Ala
213 15 20 25
214 caa gac caa tac tgt gcc ttc ata cct caa tgt aga cca aga tca gaa 269
215 Gln Asp Gln Tyr Cys Ala Phe Ile Pro Gln Cys Arg Pro Arg Ser Glu
216 30 35 40
217 ctg att aaa cct atg gat gat ata tac caa aga cca gtc gag ttt cca 317
218 Leu Ile Lys Pro Met Asp Asp Ile Tyr Gln Arg Pro Val Glu Phe Pro
219 45 50 55
220 aac ctt cca tta aaa cct agg gag gaa agcgctatga gaggatcgca 364
221 Asn Leu Pro Leu Lys Pro Arg Glu Glu
222 60 65
223 tcaccatcac catcacctgc ag atg tat aca gat tta aaa gat aaa gta gtt 416
224 Met Tyr Thr Asp Leu Lys Asp Lys Val Val
225 70 75
226 gta att aca ggt gga tca aca ggt tta gga cgc gca atg gct gtt cgt 464
227 Val Ile Thr Gly Gly Ser Thr Gly Leu Gly Arg Ala Met Ala Val Arg
228 80 85 90
229 ttc ggt caa gaa gaa gca aaa gtt gtt att aac tat tac aac aat gaa 512
230 Phe Gly Gln Glu Glu Ala Lys Val Val Ile Asn Tyr Tyr Asn Asn Glu
231 95 100 105 110
232 gaa gaa gct cta gat gcg aaa aaa gaa gta gaa gaa gca ggc gga caa 560
233 Glu Glu Ala Leu Asp Ala Lys Lys Glu Val Glu Glu Ala Gly Gly Gln
234 115 120 125
235 gca atc atc gtt caa ggc gat gta aca aaa gaa gaa gac gtt gta aat 608
236 Ala Ile Ile Val Gln Gly Asp Val Thr Lys Glu Glu Asp Val Val Asn
237 130 135 140
238 ctt gtt caa aca gct att aaa gaa ttt ggt aca tta gac gta atg att 656
239 Leu Val Gln Thr Ala Ile Lys Glu Phe Gly Thr Leu Asp Val Met Ile
240 145 150 155
241 aac aac gct ggt gtt gaa aac cca gtt cct tct cat gag cta tct cta 704
242 Asn Asn Ala Gly Val Glu Asn Pro Val Pro Ser His Glu Leu Ser Leu
243 160 165 170

```

RAW SEQUENCE LISTING ERROR SUMMARY  
PATENT APPLICATION: US/09/913,494A

DATE: 12/11/2002  
TIME: 09:37:40

Input Set : N:\Crf4\12022002\I913494.raw  
Output Set: N:\CRF4\12112002\I913494A.raw

Invalid Line Length:

The rules require that a line not exceed 72 characters in length. This includes spaces.

Seq#:1; Line(s) 2

**VERIFICATION SUMMARY**

PATENT APPLICATION: US/09/913,494A

DATE: 12/11/2002

TIME: 09:37:40

Input Set : N:\Crf4\12022002\I913494.raw

Output Set: N:\CRF4\12112002\I913494A.raw

L:6 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:450 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:12,Line#:0